An Architected Approach to Integrated Information

Federated Enterprise Data Warehousing – A Management Overview or "They Keep Moving the Cheese"*

November 2004

Dr David Waddington
Founder, Tyson Consulting BV
(Ex Chief IT Architect for Unilever Foods Europe)

Contents

Summary

Today's business environment is characterized by the challenges of increased globalization, the need to sustain growth in mature markets and the sudden explosion of legislative and regulatory compliance demands. In addition, multinational organizations are also chartered with the seemingly conflicting requirements to increase corporate agility by simplifying business processes and the IT systems landscape, while simultaneously processing and managing the vast amounts of information resulting from these many new initiatives.

Information that is timely, consistent and readily available underpins each of these business goals and is crucial to delivering on these objectives. In many large organizations, however, attempts to provide a single integrated view of all data ('a single version of the truth') have been largely unsuccessful. The ability to surmount this 'information integration challenge' is bedeviled by three enemies:

- Change: changing market demands create a constant state of flux, thus requiring business processes to remain flexible at all times.
- Diversity: lack of common information standards and the diverse nature of the information sources create a disconnect within organizations due to conflicting data.
- 'Local' vs. 'global' tensions: many large organizations are composed of autonomous business units, each with their own 'local' demands. Aligning them to corporate needs and vice versa continues to be a key challenge.

In order to cope with these integration demands, a new breed of enterprise data warehouses is required that has the intelligence and capability to manage and adapt to the challenges faced by businesses today. From a technical level, such a data warehouse would need to take an iterative approach to data warehousing, allow for rapid prototyping, be able to scale and expand to support business needs and also provide support for a common language for data across the business (master data management). From a business viewpoint, such a data warehouse would also need to address what has been unattainable to date–usability at the business level and the ability to be able to respond rapidly to business change.

Master data management enables organizations to draw on a common language for data that defines how products, customers, components and suppliers are described across all operational systems. This common set of master data is the key to successfully solving the information integration challenge.

When architecting a solution, be it a single enterprise data warehouse or a 'federated' collection of warehouses, organizations need to consider the following factors to successfully deliver integrated information:

- Company structure and business market needs: there is often a business need to retain a degree of 'local' autonomy
- The degree of divergence in the business model: there may be a need for widely different levels of detail and many different classifications of data
- The degree of divergence in the data: one part of the business may not need to know detailed information relevant to another part of the same business
- The number of sources of data: e.g. a single enterprise resource planning (ERP) system or a collection of nominally similar, but in fact different, ERP systems

Business intelligence (BI) software vendors currently offer four broad approaches to solving the information integration challenge: the custom-built enterprise data warehouse, the ERP data warehouse, the virtual data warehouse and the adaptive enterprise data warehouse.

While ERP-centric data warehouses can provide 'static' operational reporting, they fail to provide and keep up with the constantly changing, strategic, enterprise-wide performance information that companies need in order to deliver business improvements. Similarly, custom-built data warehouses struggle to stay up to date with the changing enterprise because they were built to a business model at a point in time. Virtual enterprise data warehouse approaches aim to solve real-time data integration needs by integrating and querying information from various data sources without relying on persistent data warehouse database storage–but generally aren't very scalable.

As a result, a new category of *adaptive* enterprise data warehouse solutions has emerged within the enterprise BI software market to address the shortcomings of the more traditional alternatives. It is to these vendors, such as Kalido, that we must turn for the real innovation in implementing truly integrated information.

Introduction

In his book *Who Moved My Cheese?*, Spencer Johnson tells a parable which uses searching for 'cheese' in a maze as a metaphor for finding the good things you want to achieve in life. He explains that change is here to stay, and we should be ready to change quickly, again and again. He shows us how to adapt to change quickly. If finding our 'cheese' in the maze is achieving the benefits of integrated information, then the single biggest challenge facing us is finding a way to adapt quickly, again and again, to business change. Because just when we thought we had cracked the integration challenge they 'move the cheese.' And 'they keep on moving the cheese!'

This white paper is intended for the business and IT decision maker, as well as the key influencer, who is confronting the challenge of implementing an information integration architecture for their business. While there are a number of approaches to information (data) integration, we focus here on what is probably the most commonly adopted method in Global 2000 companies today, namely the implementation of an Enterprise Data Warehouse (EDW)[1]. The business examples and cases cited are also centered upon this approach and draw heavily upon the author's past experience in both Unilever and other multinational organizations confronting the information integration challenge. The cases are generally, but not exclusively, based upon implementations featuring the KALIDO application suite (see later).

Business Context

In today's highly competitive business environment, more business is being conducted across conventional geographical boundaries, leading to increased **globalization**. Ever increasing demands from our customers require faster response times and greater fleetness of foot or **agility**. Many markets in G8 industrial countries are relatively established, and **sustained growth** is an increasing challenge for most mature businesses. Increased pressures on costs, coupled with the growth in process complexity, are forcing businesses to seek new ways to **simplify** both the business processes and the IT systems landscape. The key driver for this is closely linked with the need for increased agility and adaptability in the face of new business opportunities and increased customer/consumer demands. The recent focus on Sarbanes-Oxley in the press has increased pressure on **compliance** with legislative and regulatory requirements. Last, but by no means least, is **information**. Timely, consistent and readily available information underpins each of the previously noted business goals and is crucial to delivering on these objectives.

The business challenges outlined above are, in turn, directly coupled with the rapidly increasing pace of change in most businesses. Some of the ways in which this manifests itself are:

Business organizational change	▸ Acquisitions and disposals of business entities sometimes happen at the rate of one company per WEEK! ▸ Companies even go through major transformations such as Intelsat[2] did in 2002 when it changed from a public, government-owned organization into a private sector company. ▸ Internal restructuring such as a merger of two departments can happen in different areas of a large corporation several times a year.
New or discontinued products lines and services	▸ Increased demand for innovation in mature markets often leads to new product line introductions at a rate of one per month. However, information about discontinued products still needs to be retained to maintain their history.
Changing Customer Relationships	▸ Our customers restructure too (also as frequently as weekly!) and they are not obliged to tell us. This leads to problems in aggregating invoice information up the customer hierarchy. ▸ Market consolidation such as Unilever LA experienced when one of its customers purchased another of their customers
Market pressures	▸ New business model approaches, e.g. the recent price war in the Dutch supermarkets, have forced suppliers to quickly rethink how they do business in that market
Regulatory compliance	▸ The impact of Sarbanes-Oxley and also the Basel II risk management initiatives require the banking industry to now hold seven years of history.

This high degree of change and churn is one of the main reasons why it has proved difficult for organizations to successfully integrate their information across functional boundaries and across the enterprise. Certainly this is true of mature businesses where we seldom have a "greenfield" situation–we inevitably have to cope with the legacy of the past.

But the business drivers for integration are now increasingly compelling because Global 2000 companies must become more highly flexible than ever before to successfully compete in global markets[3]. Some examples are:

- Identifying new business opportunities by utilizing untapped information in the disparate collection of systems that reside in the organization.
- Meeting the increasing demand for central and regional views of information due to globalization
- Retaining local insight and expertise while globalizing those aspects of the business that will gain economies of scale
- More effectively managing the expanding scale and scope of the supply chain
- Increasing focus on flexibility of customer interaction and service delivery
- Integrating all customer information across the enterprise (where this makes sense) to be in a better position to meet and anticipate customer needs

- Extending reach to a broader customer base–e.g. over the Internet
- Speeding up product delivery and customer response by establishing closer systems integration with customers and vendors (requiring common data standards)
- Streamlining information flows between people and processes to eliminate delay and unnecessary work
- Obtaining faster information on the effectiveness of advertising spend

While not exhaustive, this list underlines the current need to refocus on poor information integration as a major roadblock to progress in large businesses.

In a survey by CIO Magazine[4] in March 2002, and a complementary benchmark study by IDL[5] in 2002, corporate information officers cited integration of information as their top or critical strategic priority. Unfortunately, two years later, in 2004, CIOs are still struggling to resolve this issue.

Clearly the market requires an architected information integration approach, which can handle business information the way it really is–sometimes inconsistent, sometimes incomplete. Below, we discuss the factors that need to be considered in designing such an approach, and what new packaged data warehouse technologies can offer to help architect a solution to the information integration dilemma.[6]

The Integration Challenge

In many large organizations, attempts to provide a single, integrated view of all data have been largely unsuccessful. Early custom databases, ERP systems and custom built data warehouses have failed to embrace the global business requirement for "a single version of the truth." As we noted in the previous section, however, this is not just a question of technology. There are at least five main reasons:

Information diversity and explosion

In the past decade, there has truly been an explosion in both structured and unstructured information. For example, Lyman et al has estimated that 800 megabytes of information is generated annually for every man, woman and child on earth![7] Large organizations have often grown by acquisition. From a data perspective, however, consolidation of individual businesses and information (data) from distributed and widely diverse data sources does not in turn lend itself to consolidation into a single 'corporate' or enterprise data warehouse. As we saw in the previous section, giving consistent access to such diverse data has proved a major obstacle for most businesses[8].

This has been made significantly more difficult by the absence of any form of 'common coding' or common nomenclature in such businesses. In one case, a business had twelve different definitions of NPS–all with the same name–Net Proceeds of Sales. This lack of a common shared nomenclature and taxonomy is probably the single most significant barrier to achieving consistent data. Experience shows that while everyone in a business will agree that having such a common standard makes absolute sense, gaining business-wide acceptance can prove daunting. As one who has spent 15 years pursuing this goal, I can testify to the complexity of the corporate challenge.

Change–identifying the enemy

It is folly to believe that the job of consolidating data in the business will ever end. New investments have to be made by corporations with an eye on future growth. New or reengineered business processes have to be implemented to accommodate new opportunities. Just when you think we now have a consistent standard set of definitions, the company makes

a new acquisition or changes its business operating model, bringing the process almost 'back to square one.' Change cannot be ignored: in today's business world, agility to respond to customer requirements is the difference between success and failure. That agility is contingent upon the ability of IT systems to accommodate this change at the same rate of business change. Because 'they keep moving the cheese'!

Continual change in the business makes projects difficult to deliver (an acquisition during the project can result in substantial rework). The bigger the project and the more time it takes, the more devastating the changes in the business processes can be on the project delivery. Which of us has not experienced the situation of asking for information to be aggregated differently, only to be told by the IT function that it will take three months to implement?

Technology–help or hindrance?

Early attempts to build data warehouses were disappointing. Often they were built in one go–all or nothing–frequently delivering nothing. The focus was often on the IT solution (the RDBMS) not the business problem. Too much had to be custom-built with inherent inflexibility and long delays. The problem is that if the business process changes, or the scope changes, or the data is needed for another process, not only does the database design need to change, but the code (often millions of lines of it) also has to be modified and extended. Clearly this is not a fast process–it's slow, error prone and expensive. It is interesting to note that Andreas Bitterer of META Group comments: "Sooner or later, all organizations implementing data warehouses face the problem of the data warehouse not reflecting the current state of the business. The worst-case scenario is that this goes unnoticed, and the results from query and reporting tools are simply wrong. In other cases, data warehouses are constantly remodeled to the point of being unusable because of continuous patchwork of the structure."[9]

The result is that businesses often cobble together solutions to meet an urgent business need, only to find at a later stage that they are limited by problems with flexibility, availability and scalability. More seriously, new opportunities are frequently missed because the nuggets of information required for making business decisions get lost in the focus on technology.

Various vendors have launched a plethora of tools–all proclaiming to offer an overall solution–but the bottom line is that customers often find themselves integrating the "integration solutions." While some vendors will pre-assemble the solution with 'best practice' models and pre-built integration code, but these tools lack the ability to truly address the challenges mentioned above. In truth, instead of helping, these "solutions" just help the business sink further into the mire of IT complexity.

Tension between "local" and "global" requirements

Most large organizations are made up of a number of relatively autonomous operations. In the Royal Dutch/Shell Group of Companies (Shell), for example, there are over 1000 separate operating companies ranging from just a few employees to tens of thousands of employees per business unit. The group changes through reorganization, acquisition and divestment on a regular basis. Business priorities (and hence information needs) are very different from one part of the organization to another. Much the same is true of Unilever, which has operations in some 80 countries handled by more than 350 separate operating companies with coordination at the Business Group level.

For example, a small local company within a business group may wish to set up a limited operation selling specialized cleaning services utilizing their products. They need to act fast to secure a position in an untapped market. A global category director with responsibility for managing and optimizing his or her category (say industrial cleaners) may need to know the profitability and market share of each product by country. Both would benefit from having shared, integrated business information but the local operation needs to move now while the category director understands that it may take quite some time to deliver the full range of information he needs.

The small operation has a culture of 'going it alone,' while the enterprise as a whole has always juggled a 'loose federation' of interdependent businesses. The enterprise has recently initiated a program to 'globalize' the corporation.

Clearly, very different priorities, very different needs and cultures. This tension between 'local' and 'global' needs makes for a difficult marriage and involves difficult 'project politics.' It is sometimes suggested that a more radical approach should be adopted in such circumstances; that the individual businesses must be made to 'toe the line' of commonality. While this may not be wrong in principle, in an organization with a deep-seated individualistic culture, this will take time to change (all too often underestimated), and time is not always available. Furthermore, companies should take care to avoid forcing conformance to standard processes so that opportunities afforded by a more flexible approach are not lost.

In the politics of project management, balancing conflicting priorities is a serious problem. When one set of priorities is sacrificed in favor of another, the organization as a whole often comes off worse. We must recognize that different needs and priorities exist. None should be considered wrong. All must be able to co-exist if the business requires this.

Large projects are inherently risky

The very thought of trying to marshal senior business representatives to decide and agree upon strategic, global priorities and standards for building a global enterprise data warehouse is daunting. Coupling that with the risks associated with such an undertaking means that it is often not even attempted. Two key reasons that very large projects struggle to deliver value is that they grind slowly to a halt due to large amounts of re-work necessary to adapt to new business requirements, or resource requirements are increased to a point where the cost is too high to bear.

One solution to this problem has been to break down the project by building a collection of separate data warehouses and then trying to link these to form a loose 'federation.' The clear advantage is that this, in principle, reduces the magnitude of the problem and allows a step-wise or phased approach to the project. A similar approach, adopted by many organizations to try to circumvent these problems, has been to build a number of 'data marts' (subsets of 'cleaned' data), often encouraged by BI tool vendors or systems integrators. The reality is, however, that most of these companies have ended up with subject-oriented 'warehouses' (really data marts) that are not interlinked and which have led to immense data duplication and high maintenance costs.

Such approaches have further been hampered in part by the absence of a framework for phasing such an ambitious implementation. The small number of technically skilled people available to build software capable of performing under such imposing conditions means high levels of risk. Coupled with that, the limited availability of business representatives who have the executive authority to agree on priorities with the rest of an ever-changing organization is enough to put global integration project managers on the endangered species list.

Without the technical architecture of an 'off-the-shelf' product that allows local modifications and a coherent global view to coexist, it would be a very brave project manager indeed that committed to delivery on time and within (any) budget.

What is an Enterprise Data Warehouse?

An enterprise data warehouse (EDW) is an implementation of a holistic model of the business which includes such items or objects of importance to the business as customer, product, time, geography, sales hierarchy and market (sometimes referred to as 'dimensions' since they define the context of the business transactions). It is a database in which atomic level data from disparate sources is brought together in a structured way creating one multi-subject oriented version of the corporate truth, designed to enable timely, accurate decision making in support of strategic and tactical business initiatives.

As outlined previously, there are basically four different approaches (custom-built EDW, ERP data warehouse, virtual data warehouse and the new Adaptive EDW) to implementing an enterprise data warehouse.

Custom-Built Enterprise Data Warehouse

As the name implies, custom-built implementations are specifically designed for a given business model and often inflexible to change.

Adaptive Enterprise Data Warehouse

A new breed of data warehouse, the adaptive data warehouse has recently been introduced. This kind of warehouse is readily adaptable to changes in the business model (see later).

ERP Data Warehouse

Many ERP vendors now include a packaged data warehouse in their product suite, but these are usually focused around operational reporting from the ERP package. They are most often 'custom-built' data warehouses designed by the vendor for their ERP package.

Virtual Data Warehouse

A number of approaches have used advanced messaging technologies (e.g. Enterprise Application Integration) to link together disparate data sources into a 'virtual' data warehouse

Key Requirements for any Enterprise Data Warehouse

These are the key factors that have made, and continue to make, it difficult to deliver integrated information in large organizations. What is required to break out of this dilemma? Viewed from my previous role as Chief IT Architect, analysis of the above challenges leads me to the conclusion that any enterprise data warehousing solution must be able to encompass the following features:

Speed and Flexibility

- It must offer an **iterative approach** for implementation to allow changing business requirements to be taken into account. Allowing a solution to be provided quickly for a specific business area while not restricting future extensions.
- It must enable **rapid prototyping** so that a new business requirement can be quickly explored together with business users.
- In other words, the business model represented in the warehouse should not have to be 'hard-wired' in advance. Often people realize that they need different information from the warehouse than they first thought.

Scalability

- The approach must be **scaleable** such that the overall implementation can be split into a series of separate (smaller) projects to compensate for the different priorities held by different parts of the organization. This will also simplify the overall implementation by avoiding the need to run very large projects. Populating the master data (see later) is essential to making it possible to aggregate data to deliver a consistent view in this situation.
- Ideally the warehouse should be easily **extendable** to permit inclusion of new business areas or functions or new sets of transaction data. For example, it should be possible to build a data warehouse first to tackle the Sales and Marketing area in a business and then extend it to include Finance.

Expandability

- As the business develops, grows and undergoes change, potential new sources of data will undoubtedly be identified (e.g. an acquisition). It is essential to be able to incorporate these **new data sources** quickly into the data warehouse without the need for complex and time consuming redesign.

User Oriented

- Frequently we would expect to use the data warehouse as a coherent source of data for the creation of various **subsets of data** to facilitate reporting such as **data marts** or **data cubes**. The data warehouse should support the process of creating these data marts and assist the user in extracting valid sets of data.
- The **design approach** for the data warehouse must enable the business users to be intimately involved with the initial design and subsequent design changes to meet the changed business needs. Experience shows that the greater the involvement of the business users in the overall design of the data warehouse, the higher the chance that it truly meets real world business needs. For example, it must be able to cope with the fact that often a business user may refer at one point to a 'customer,' which in other circumstances may be viewed as a 'supplier.'
- Ideally, the data warehouse must be adaptable, and the users should only need to operate at the **business level**. They should not need knowledge of the names of the tables in the system. Equally, it should not be necessary to unload all the data from the warehouse and then reload the data again following changes to the warehouse made to reflect changes in the business. This slows down development and hampers the ability to respond quickly to business change.

Respond to Change and Business Modeling

- The data warehouse must have the ability to cope with **time dependence** and variance–i.e.: to support the notion of 'corporate memory,' the business past, present and future. Organizations want to build data warehouses to be operational over long periods of time to enable the capture of historical information. During this time some objects will change. Unless this 'history' can be captured, the data warehouse will not be able to support the business requirement to compare like with like over time.
- This is essential to support business **scenario planning**, the functionality to view and analyze data in a variety of ways including historical trends so as to be able to explore the implications of potential future business options and opportunities. Furthermore, new reporting scenarios and structures are often introduced (an Annual Plan format for the coming year for example) and need to be included in the warehouse prior to them becoming valid (the 'future dates' concept–i.e.: information inserted now into the data warehouse which does not become valid until some date in the future). This too is important for scenario planning.

Master Data Management

- The management of **master (reference) data**[10] is crucial. Master Data is data about products, customers, vendors, the organization, geography etc. This focuses on managing the taxonomy of common business definitions and descriptors.
- Master data should ideally be stored separately from transaction data to allow for differing coding structures in disparate source systems and more easily accommodate change in the master data (ie: recognize that master data is not static).
- When it is properly and consistently managed, master data provides a consistent context within which business performance can be measured and monitored. It enables management of the link between disparate definitions, aggregation hierarchies and mappings.
- In the absence of well-managed master data, data warehouses deliver unreliable results at the reporting stage, which reduces their credibility and value to the business. Also, as we noted above, master data management is key to the successful design and operation of a collection of linked data warehouses.

Taken together, these features describe the functionality required for any enterprise data warehouse product. New technologies are now available which are able to meet these demanding requirements, and we will return to some of these issues in the next session and again later when we look briefly at the available technology.

This potentially takes us into a requirement for **federation**–the ability to interlink several smaller data warehouses into a larger holistic warehouse that doesn't duplicate information.

Integration using Federation

A perfectly viable architecture for a data warehouse to integrate information across your business might well be to implement a single enterprise data warehouse. There are many businesses that have chosen this route, and the choice is dictated purely by the business requirement. It's certainly a good starting point for any analysis to determine the best architecture for you to employ in your business. We mentioned earlier that there have been many attempts to build or interlink collections of smaller subject, focused data warehouses (each built to the same data model), but that these have generally been unsuccessful. This is generally referred to as 'federation'[11] and is illustrated in Figure 1 below.

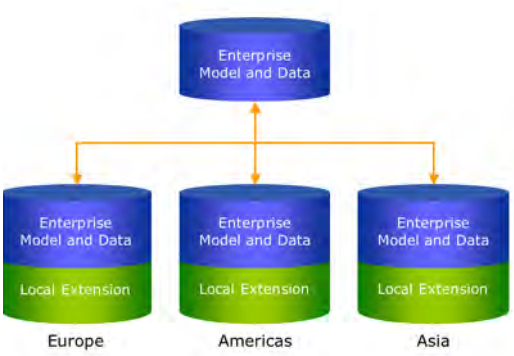


Figure 1
A Federated Architecture

Most large companies' organizational structures balance the need for local autonomy and flexibility with the need for standardization and central control. Probably this presents the greatest integration challenge since this requirement cannot effectively be met by a single data warehouse implementation.

A **federated architecture** reflects the organization structure by dividing authority or 'ownership' between a central entity and its semi-autonomous constituents or 'federation members.'

It aims to strike a balance between two extremes: at one extreme is to build a comprehensive corporate data warehouse that serves all regions and business areas; at the other are separate data warehouses for each region or business areas with complete autonomy. Alas, all too often a corporate warehouse is built which ONLY serves the needs of the center, e.g. a financial consolidation tool becomes the de-facto enterprise data warehouse. The federated model for data warehousing is one in which a hierarchy of data warehouses can exchange data, business models and reporting structures, to allow local autonomy and customization, but can also deliver global control and a degree of standardization. A 'federated data warehouse' consists of a set of data warehouse instances that operate semi-autonomously, are generally geographically or organizationally disparate, but which can be thought of as one large data warehouse. Of course a single enterprise data warehouse can sometimes be built where the business is uniform (e.g. Wal-Mart) but even this can be destroyed when an acquisition occurs (e.g. Asda[12]).

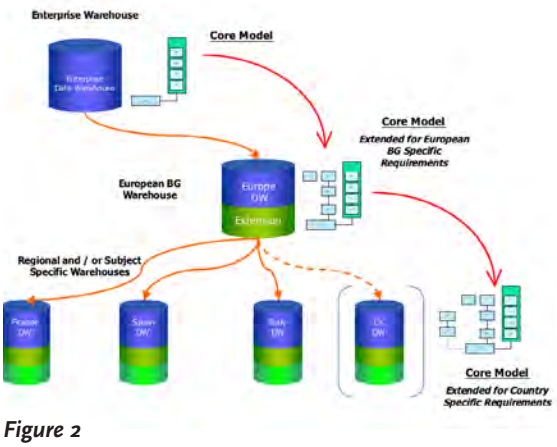


Figure 2
Transfer of a Core model

So for example, at the corporate level, the high-level product structure (classification) for a given business area, say Impulse Ice Cream, could be defined and passed to the European Business Group data warehouse as a standard (see Figure 2).

The European business group is free to extend this with its specific European view of products in Europe while maintaining the corporate view intact. In turn the European business group can pass this to the local level together with the appropriate extensions. The local operating business could then extend the product classification (or add additional local classifications) while maintaining the corporate and regional views. It's worth noting here that with conventional data warehouses this would mean destruction of the data model.

Since a federated data warehouse can be built one step at a time, it offers a "start small, think big" approach to enterprise data warehousing. The federated approach can significantly reduce risk in a global rollout, because each local warehouse is smaller in scope, delivers quickly on local requirements, and can be operated by local business units.[13]

An Integration Framework

So, a 'federated' enterprise data warehouse architecture offers a promising solution for many large organizations with a level of internal autonomy to the integration challenge.

This section offers a framework for analysis, and outlines some of the key factors to be considered based upon experience in deciding whether adopting a federated data warehouse approach is appropriate in your business, and if so, the optimal architecture for such an endeavor[14].

To Federate or Not to Federate?

What is involved in designing and building the framework that will support information integration based upon data warehousing? Be clear at the outset that there is no single "correct" answer. For some businesses, a single data warehouse may be appropriate (possibly to be federated at a later stage), whereas for others, a federated model may be the ideal starting point. It is important to be clear at an early stage which approach is appropriate for your business.

The key factors that need to be considered are outlined below. They are drawn from experience gained in considering the advantages and disadvantages of the potential options outlined below within Unilever, Shell, BP, Philips and other major organizations.

Company Organization and Business Requirements

One reason for deciding for federation may be indirectly linked to the organizational structure. However, a hierarchical internal structure does not imply that federation is automatically the best solution. The key factor is the degree of delegated or local autonomy. In my experience, this is closely linked to the level of 'local data ownership.' If local users are used to owning and executing changes locally then federation is a good option. It also brings the added advantage of obtaining 'buy in' from those who feel threatened by a central group 'owning' their data.

Experience also shows that when the organizational structure involves a high degree of local autonomy (which may be desirable for some businesses), it will just be a question of time before all the requirements, which were initially agreed to be the same, diverge. This situation cannot be satisfied with a single instance approach and underlines the fact that it is essential from the start to adopt an architecture that can accommodate this.

The Treacy and Wiersma model[15] states that businesses excel either as innovators, by being intimate with customers or by being efficient. It is worth noting that two out of the three (the first two) require local autonomy.

Degree of Divergence of the Business Data Model

The most important criteria for considering federation as an approach to information integration is the degree to which the various business data models are (or need to be) different. If, for example, two parts of an organization require different product classifications, it may well be appropriate to put both classifications into a single model. BP Lubricants has more than 20 product classification structures (brand, pack, usage, etc) within a single model. However, there are limits to this. If the number of these small differences is likely to become large and systemic, then the final single instance approach would be unmanageable. Experience shows that this has often been the cause of the failure of conventional custom-built data warehousing projects.

Major structural divergence is another strong reason for federation. If, for example, a foods service business sells directly to consumers, while another business unit sells only via distributors to customers, then the transactions will have very different contexts and structures (classifications) making it difficult to accommodate both in a single business data model.

Degree of Divergence of Data Content

Often one business unit is not dependent upon data from another to make decisions relating to its business. In Unilever, for example, Unilever Foods in North America generally does not need to see the details of sales to all customers in Europe in order to decide its sales strategy for the next month. There may, however, be some common master data, for example a common corporate classification of customers that is used to classify the individual customers. If it is clear that each potential federation member needs all transaction (and master) data, then a single warehouse is the better option.

Number of Sources of Data

Data sources for the content are also very important considerations. If there is one main source for all data, having to filter the data to load separate warehouses would result in extra work. If data are sourced locally, federation becomes more viable.

The decision to opt for a federated approach to data integration needs to be taken based on an analysis of these considerations. The general 'rule of thumb' is that the greater the degree of autonomy of business units across your business, the more appropriate it will be to select a federated option.

The Importance of Master Data Management

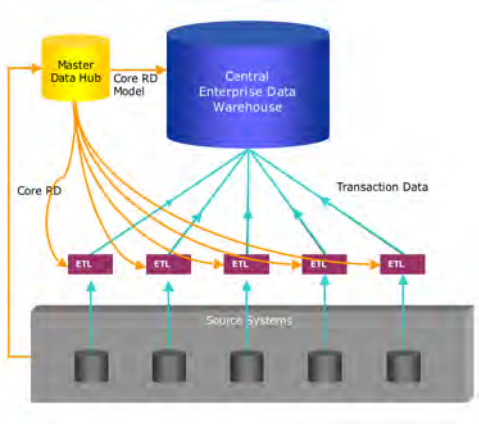


Figure 3
Master Data and the Single Instance Data Warehouse

There are two aspects to master data management relating to process and to data modeling. In the present context, it is appropriate to focus more on the process; however, both are key to effective master data management[16]. In a sense, master (reference) data management can be thought of as a special form of data warehouse federation. The key concepts are the same: reference data, such as product, customer, vendor and organization structure, are stored in a master data warehouse. This warehouse contains core product information needed by all the systems (federation members) that subscribe to it, but the subscribing systems would take the core product information and extend it: marketing may create multiple 'local' products per single corporate product classification; manufacturing may add bill-of-material information.

Before turning to Master Data Warehouse architectures for federated warehouses, it is important to emphasize that effective master data management is also key to the integration of data from several sources into a single data warehouse instance.

This is illustrated in Figure 3, which shows a single instance enterprise data warehouse which is being loaded with transaction data from a number of disparate sources using ETL[17] (extract, transform and load) processes.

During the loading process, master data from a Master Data Warehouse is merged to ensure consistent mapping of transaction data to bring this in line with the data warehouse business model.

The key issue regarding effective implementation of integrated data within a federation is to ensure that the transactional data flowing between the federation members is always based

on common consistent master data. This can best be achieved by implementation of a Master Data Warehouse. It is in principle possible to build a federated model that does not rely upon the use of a Master Data Warehouse but instead employs message-based routing and transformation technologies (EAI) to transform data en route.

However, experience shows that such systems are complex to manage and that, over time, the absence of a common warehouse (a single common source of master data) leads to inconsistency and frequent errors in transactional data. Furthermore, such systems generally do not easily lend themselves to being rapidly adapted to new business requirements since they are often dependent upon writing low-level code to effect the revisions.

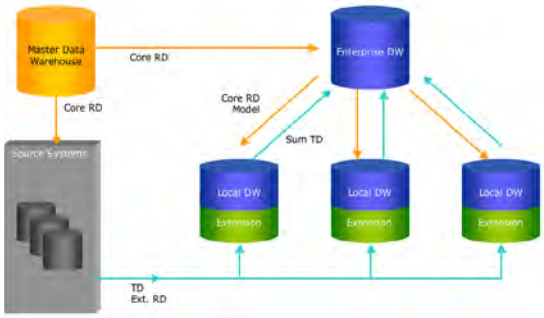


Figure 4
A Master Data Warehouse

The high level architecture for a federated set of data warehouses with integrated data, based on maintenance of master data using a Master Data Warehouse, is illustrated in Figure 4. Here the master data warehouse supplies master data to both the data warehouse federation and the underlying source systems.

It is important to note here that the Master Data Warehouse is in effect an **operational system** since it manages all master data flowing through the organization.

This demands effective processes and procedures for management and ownership of master data.

While the above ideally should be the longer-term goal for the organization, an excellent first step in this direction can be taken by using a Master Data Hub as illustrated in Figure 5. The source data will inevitably need augmentation, reviewing and/or mapping if it is to meet the project's needs.

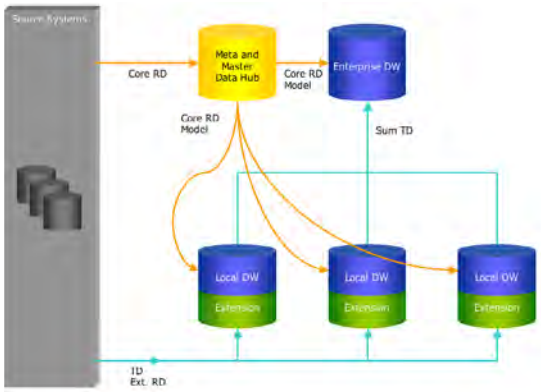


Figure 5
A Master Data Hub

The advantages of this approach are that the metadata (e.g. the core data model) and master data are distributed from a hub rather than from a central data warehouse or source system. Thus consistency can be assured while this approach can be very flexible in addressing variation and differences in the business data model and divergence. The hub approach supports the idea that not all master data needs to be created in the hub, and can also be created in designated applications.

This approach has essentially been used by Philips[18], where a separate warehouse (one not holding transaction data but only master data) is used to manage and publish master data to a series of federated warehouses.

An increasing number of large businesses including InBev, Philips, BP and Unilever are recognizing that without common managed master data, integration of information is difficult or impossible to achieve. These companies are now taking the first steps to put in place a managed master data warehouse that will manage master data across their businesses.

Some Architectural Options

There are many architectural alternatives for integration using a federation design; however, the overall design must be based upon the requirements of the business organization and process as outlined above.

Broadly, there are three types of federations: **geographical**, **business area** and **functional**. In a geographical federation, the members reflect the business needs of each geographical region, e.g. North America, Asia, Europe or even individual countries.

Business area implies that the subject focus of each member of the federation is a line of business, e.g. in Unilever that could be Home and Personal Care, Foods or even Ice Cream and Frozen Foods. A functional federation approach is focused on the business functions, such as Sales, Finance, Marketing and Supply Chain. Conceptually the three are similar, though the actual design can differ widely since the delegation of authority works very differently in each of these three scenarios.

Additionally, there is the option of using a single physical data warehouse.

Single Physical Enterprise Data Warehouse

At first sight the option of using a single data warehouse is appealing (see Figure 6 below). Unilever Latin America, HBOS plc, Cadbury Schweppes[19] and Intelsat have all adopted this approach with significant success (details of these case studies can be found on www.kalido.com). In line with the considerations outlined above, Unilever LA selected the single data warehouse approach to underpin their business objective of moving toward a truly regional operating model with regionally common processes. However, it is believed that they will eventually be a federation member of Unilever global.

Furthermore, data is sourced from a single regional SAP R/3 ERP system together with other single instances of systems such as Siebel and PeopleSoft. This would have made extraction of data to a potentially large number of 'local' data warehouses costly and complex.

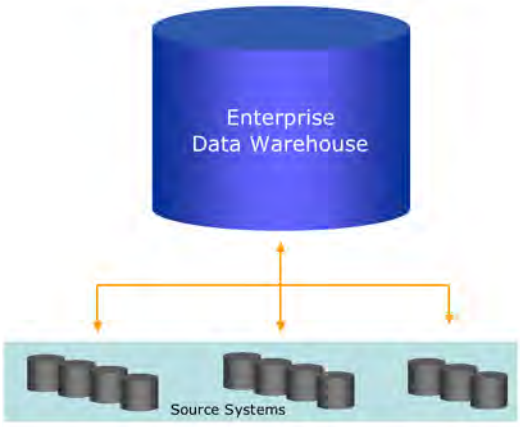


Figure 6
Single Enterprise Data Warehouse

Their desire was to move away from this 'local approach' (they originally had some 34 custom built local data warehouses). A key requirement was to be able to generate a number of specific data marts from the base warehouse data focused on different business functions. This is a typical case where, on balance, opting for a non-federated solution is to be preferred. One advantage of this architecture is that if, in the future, there are sound business reasons to move to a federated model, then this remains a viable option.

Intelsat has noted that while they have initially (given their size and business requirements) opted for a single instance data warehouse, merger with or acquisition of another major company would lead them to adopt the federation route– retaining their current warehouse but federating this with the new business.

As indicated in the previous section, the single large data warehouse option really works best where there is little room for autonomy and where there are strong management principles for incorporating local extensions. Generally in Unilever LA, extensions have been incorporated only when there is a regional need (i.e. more companies require the same or similar functionality). On completion of their current program, Unilever LA expects their enterprise data warehouse to reach 10 terabytes in size.

The downsides to this approach include performance lag and throughput. How do we produce a reliable system that can maintain very large dimensions, while simultaneously supporting fluctuating demands for data access? How do we balance the needs of indeterminate ad-hoc query access with the maintenance of ever changing reference data? How do we provide systems with enough raw computing power to service such demands?

Techniques have emerged to address these problems aimed at increasing the raw power available to throw at the situation. However, the cost of resourcing such systems is high. Secondly, they pose a risk: the entire business is dependent upon the resilience of a single warehouse. One final, but highly undesirable, feature of large monolithic systems is that they have to be tuned. Not easy when demands are coming from all over a diverse organization and are sometimes in direct conflict with one another. Clever aggregation strategies and mart design can help to mitigate this, but, tuning a system to try to meet ALL the requirements of a large organization is a major challenge.

Geographical, Business Area and Functional Area Federations

Governed by the criteria outlined previously, the alternative way to achieve information integration is to federate broadly based upon one of the following three models.

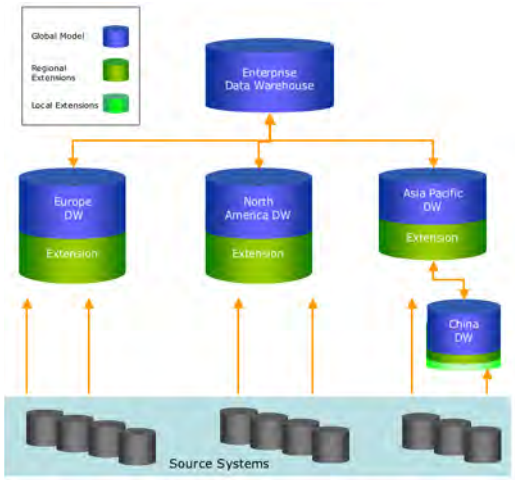


Figure 7
Federation by Geography

Essentially they can all be represented by variants of the diagram shown in Figure 7, which illustrates a geographical federation architecture.

This shows a global business, such as Shell, with a series of regional data warehouses as second tier federation members. Optionally, for large countries there can be a first tier federation with the local country operating business units. At the end of 2003 Shell Oil Products had over 70 separate data warehouses serving its operations in 80 countries. They maintain some common master data (products, customers, etc) in a central core model, which then gets distributed out to the regional and local installations at regular intervals. The common master data is added to and modified locally if appropriate.

A somewhat similar approach has been adopted by Unilever in Asia. In this instance, the focus has been to use a loose federation to collect and aggregate sales data over the region on a per country basis.

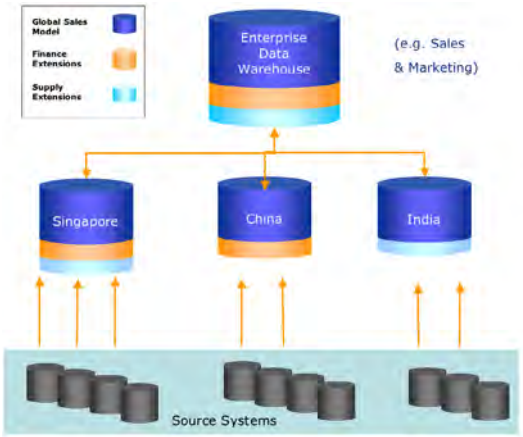


Figure 8
Federation by Business Functional Area (e.g. Sales + Marketing)

For Unilever in Asia, the approach was to opt for federation based on business function (Sales & Marketing) because this had the highest business priority.

This is illustrated in generalized form in Figure 8. The federation can be extended with other business functional areas, such as Finance and Supply Chain, as appropriate. Here again, the approach has been to build a standard model with a standard core set of master data and the opportunity for local extensions.

If the data warehouse for a single country or functional area becomes unavailable for some reason (serious though this is), this represents only a part of the federation and can be more easily accommodated than the loss of an entire information system.

If we sell part of the business, then we do not have to cope with 'unraveling' this from a single enterprise data warehouse, but can instead extract the relevant information or even transfer the local data warehouse instance. While it is perfectly possible to run a single production data warehouse (a number of companies are already doing so, such as Unilever LA, Intelsat, Owens Corning, HBOS plc.), introducing federation can help to split the risks removing the dependence for continuity of operations from a single hardware platform.

Other examples of full or partial federated approaches have been implemented by Shell, BP, Philips and Unilever (details can be found on www.kalido.com).

Examples of full or partial federated approaches

Shell Oil Products MIS (Management Information System)

Shell OP needed to segment global customer, product and channel information to support and monitor global marketing initiatives. Shell OP also needed to standardize management information without imposing new structures on local operating units, which have local data models and systems.

As cited earlier, they maintain some common data in a central 'core' which is then distributed out to 'satellite' installations at regular intervals. This common master data is added to and modified locally if appropriate, but is refreshed regularly as it changes (without upsetting local additions and modifications of course). This is a fully federated geographical approach and has formed the basis for extending the original MIS to encompass further business areas.

This has afforded benefits such as: delivering consistent global segmentation; accommodating local variations; gaining deeper insight into global customers and products; improved measurement of global marketing performance; providing a full breakdown of sales information by product and by country; delivering faster, more efficient collection and analysis of data.

BP Business Intelligence and Global Standards (BIGS)

In strategic terms, BP Lubricants is concentrating upon further improving its customer focus and increasing its effectiveness in automotive markets. Following recent merger activity, the company is undergoing transformation to become more effective and agile, in order to seize opportunities for rapid growth. To do so, it needs to supply business managers with fast access to reliable global business information.

BP Lubricants designed the BIGS program to provide managers with globally consistent business intelligence and standardized performance indicators on a timely basis. BIGS spans four major areas of business functionality–Sales, Marketing, Finance and Supply Chain. The BIGS initiative is designed to dramatically increase the efficiency and speed of business reporting by giving BP Lubricants managers a global source of reliable data via web-based self-service reporting formats. As well as delivering global views of information, BIGS will also provide drill-down views at the country, business unit or functional level. This way, the system enables global reporting while allowing for diverse local reporting needs, such as required by government regulations.

BP Lubricants ran a successful pilot program in Greece to demonstrate the BIGS program capabilities and is now focused on moving toward a global roll out of the approach based upon a federated model.

Philips CE

Philips Consumer Electronics has a large number of business units each having different information needs. The information sources include four SAP systems, i2 for supply chain and a series of legacy systems and external data sources. The challenge was to create consistent reporting covering the range of reporting needs.

Philips decided not to build a single unified data warehouse. They felt that a single data model could only be created once the business processes had been converged to a common set. Philips believed that this would take some years to achieve.

Moving to a shared service model, Philips CE looked to maximize the efficiency of data warehousing across its organization and improve its ability to respond to change. Philips used a federated approach for master data with master data sourced from a separate warehouse to both ensure consistency and flexibility to cope with changing business requirements.

Unilever Information Program (UIP)

Unilever needed to increase the accuracy and consistency of strategic global information on financials, customers, brands and key materials, in order to meet targets for cost reduction and increased efficiency.

UIP has been implemented based upon a partial federated approach with a common core business data model and a combination of direct data extracts from other federated warehouses together with other controlled data collection systems. Key to securing the required information consistency has been the implementation of a managed master reference data warehouse. This partial federation approach was chosen to allow those business units that were already implementing a data warehouse to feed data directly from this while other areas could adopt a conventional data collection approach.

UIP has significantly improved the quality and consistency of financial, brand and customer information, delivering greater insight into Unilever's international business. Unilever now benefits from greater visibility on key materials and suppliers in its global supply chain, which provides opportunities for cost savings. The company also has faster, more detailed information on global customers, broken down by brand and by country, which will enable improvements in brand management and enhance Unilever's competitive edge.

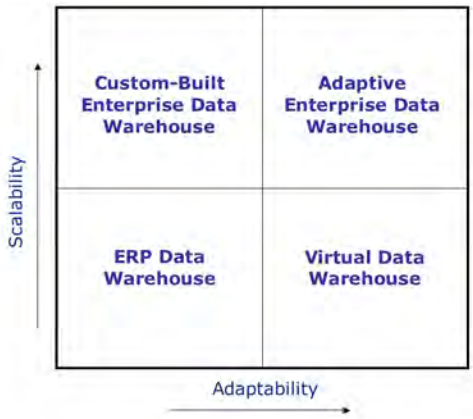
The options for architecting these federations are all dependent upon the strategy to be adopted for the management of the master data. Master data management is central to the success of a federated implementation.

Technical Solutions

Experience in the current market reveals that there are a number of technical enterprise data warehousing solutions available for the integration of information. These include:

- Solutions based on proprietary enterprise resource planning (ERP) 'packaged' data warehouses (e.g. SAP Business Warehouse®, Peoplesoft Enterprise Warehouse® and Siebel Analytics® and eBusiness Data Warehouse®),
- Custom-built data warehouse solutions from the major database (e.g. Oracle and IBM) and specialist data warehouse database vendors (e.g. Netezza),
- Virtual data warehouse[20] solutions based around messaging infrastructures and semantic analysis (e.g. IBM Federated DB2), and
- A newly emerging breed of 'packaged' adaptive enterprise data warehouses, of which KALIDO 8[21] is the only currently available commercial example.

Figure 9
Scalability vs. Adaptability Matrix



Unfortunately, there are no simple rules for selecting the 'right' solution. The optimal solution, as so often, depends upon your specific business requirements.

Experience has shown, however, that a good practical guide to selecting the best solution is based upon the volume of transactions and the amount of business change which is to be expected.

A broad positioning of the four main approaches to enterprise data warehousing described earlier is shown in Figure 9. This matrix[22] serves to illustrate the relative scalability (transaction volume) versus adaptability (ability to accommodate change) of some of the current approaches.

Technology Options

Custom-Built Enterprise Data Warehouse

- So if you have very high transaction volumes (e.g. analysis of EPOS data) you may do well to look at products in the custom-built EDW quadrant (upper left-hand quadrant).

ERP Data Warehouse

- If all of your data is held in a single ERP system, and if you do not expect much change in your business model over the next five years, then a product in the ERP DW quadrant may be the most appropriate choice. Your major requirement here needs to be focused on operational reporting.

Virtual Data Warehouse

- In general virtual data warehouse approaches aim to solve real-time data integration needs by integrating and querying information from various data sources without relying on persistent data warehouse database storage. These solutions generally aren't as scalable as solutions with persistent data storage and lack the ability to do large-scale historic analyses and future scenario planning.

Adaptive Enterprise Data Warehouse

- If, however, your requirement is to build a phased enterprise integration solution based on an enterprise data warehouse, and your business is likely to be subject to substantial change over the coming years, then in my experience only KALIDO 8 is capable of addressing the information integration challenge. This would also be the case where your data is sourced from a number of ERP systems. Even when these are based upon the same vendor ERP package, it does not necessarily follow that they will have used the same definitions and ERP model.

- The tradeoff for this flexibility of the adaptive approach is in processing time (due to generation of star schemas), which means that real time (to the second) data warehousing is not practical–but then the majority of businesses do not require this. Daily or hourly updates are usually more than sufficient.
- In addition, the technology options in support of Enterprise Data Warehousing are:

Master Data Management

- In the area of Master Data management and warehousing there are effectively only two major options at the present time, SAP MDM and KALIDO MDM. If your IT landscape is dominated by SAP then it makes sense to consider the SAP MDM option. However, in a mixed landscape with **disparate** information sources and more than one ERP solution, then KALIDO 8's master data warehousing solution is more likely to provide the flexibility required to meet your needs.

Federation Management

- It is however worth noting here that, like iterative implementation, experience has shown that federation is almost impossible to achieve using traditional custom-built data warehouse development approaches. Without going into the technical details here, KALIDO 8 offers the technology to support a federation of distributed data warehouses in order to provide a global view. In particular, KALIDO 8 provides extensive support for federation, including key functionality such as integrated Master Data Management, as well as the managed transport of business models, reference data and transaction data between federation members.

Further details about the principles that underpin the ability of KALIDO 8 to implement a federated solution can be found in the White Paper by Cliff Longman, "Data Warehouse Lifecycle Management Concepts and Principles."

By separating global and local needs, individual instances can be tuned to meet specific (and highly variable) needs meeting extreme demands of scalability. However, the package properties of KALIDO 8 allow you to easily rollout a new instance, by using its packaged meta data transport and upgrade utilities. This allows a firm degree of central control and economic administration, without compromising local autonomy. It allows the existing implementation to evolve and adapt to changing business requirements, global and local, while easily adopting new sites expanding the coverage towards global.

Finally, KALIDO 8 provides both hooks to standard management tools, as well as diagnostic and management tools of its own allowing the efficient, remote management of large collections of instances.

The introduction of the new *adaptive* enterprise data warehouse technology heralds a new era in implementing enterprise data warehouses and at long last presents a practical business oriented approach to resolving the information integration challenge. And why should we worry if 'they keep moving the cheese'–we can adapt!

Endnotes

*From "Who Moved My Cheese," by Spencer Johnson, Random House.

[1]The term 'Enterprise Data Warehouse' is defined in a subsequent section.

[2]Intelsat is a Washington DC based company, see www.intelsat.com

[3]T. Evgeniou, 'Building the Adaptive Enterprise,' INSEAD Report, www.insead.edu, 2002

[4]CIO Magazine, The State of the CIO, IT Spending, www.cio.com/state, March 2002

[5]Industry Direct Limited (IDL), 2002 Information Integration Benchmark Study, "Sharing Data Effectively Enterprise-Wide," www.idlworldwide.com

[6]C. Steensboe, 'Information Integration in the Global Enterprise,' IDC White paper, www.idc.com

[7]P. Lyman, H. Varian, A. Dunn, A. Strygin and K. Swearingen, "How much information?," University of California, Berkeley, 2003

[8]W. H. Inmon and D. Meers, "Accelerated Enterprise Data Warehouse Profitability," A White Paper available on BILLINMON.COM

[9]A. Bitterer, Approaches for Next-Generation Data Warehousing Application Delivery Strategies, META Group, 23rd December 2002

[10]The terms 'master reference data' and 'master data' are used interchangeably and have the same meaning in the context of the above. Both terms are found in the literature.

[11]A. Hayler, 'Report Global *and* Report Local,' on BusinessIntelligence.com

[12]Asda is a UK based food and clothing superstore and subsidiary of Wal-Mart, see www.asda.co.uk

[13]C. Longman, White Paper entitled "KALIDO DWLM Concepts and Principles," see www.kalido.com

[14]W. S. Chen, 'Analysis and Design of a Federation of Data Warehouses using Kalido DIW,' internal Kalido white paper, March 2003

[15]M. Treacy and F. Wiersma, '*The Discipline of Market Leaders: Choose Your Customers, Narrow Your Focus, Dominate Your Market*,' Addison-Wesley, 1995

[16]H. D. Morris and D. Vesset, 'Managing the Changing Context for Business Transactions,' IDC, www.idc.com, April, 2004

[17]See for example G. Powell, 'Best-Practice ETL for KALIDO® 8 using Ascential DataStage™,' a white paper on www.kalido.com

[18]Philips is a Dutch based international electronics company, see: www.philips.com

[19]Cadbury Schweppes is a UK based major international confectionery and beverages company, see: www.cadburyschweppes.com

[20]Sometimes referred to as EII solutions.

[21]Kalido has recently launched KALIDO 8, which offers adaptive enterprise data warehousing solutions for master data management, project-based data warehousing and federated data warehousing.

[22]Matrix based on Kalido desk research.

About the Author

The former Chief IT Systems Architect for Unilever's two Food Groups in Europe, **Dr Dave Waddington**, is one of an increasing number of IT directors who joined the field from a business background. He joined Unilever in 1974, where he led an international team working on the development of low calorie spreads and was involved in process automation. He also served as Section Head of the Linear Programming Group and was a founding member of the IT team that supported the move to open systems and standards.

Since the mid-1990s, he has worked on establishing common data standards and data warehousing and was the first to introduce KALIDO® software into Unilever. More recently, he has led a team to develop a Master Reference Data Repository, which is now being implemented in Unilever. Before Unilever, he had an academic career at the Universities of Leicester and Oxford where much of his early work focused on developing computational mathematical models for aqueous and organic mixtures.

Following his retirement last year, he now leads his own consultancy, Tyson Consulting b.v. based in The Netherlands, which offers advisory and consulting services in the areas of business intelligence, data warehousing and information management strategies.

This paper represents Dave's views, drawing upon his many years of experience in the field of enterprise data warehousing and master data management acquired both in his former business and from his extensive contacts with other multinational businesses ranging from Oil and Gas through CPG to Telecommunications.

The generous help and guidance of many colleagues both past and present is gratefully acknowledged. The author wishes to express his especial thanks to Winston S. Chen of the Kalido Competence Center whose work on the underlying design criteria for a federation of Kalido data warehouses formed the basis for the sections on 'federation' and 'architectural options.'

About Kalido

Kalido provides adaptive enterprise data warehousing software to Global 2000 companies. The KALIDO® application suite (KALIDO) delivers consolidated views of enterprise performance and can rapidly adapt them to major changes in the business such as mergers and acquisitions, reorganization, market consolidation, or new regulatory requirements. This improves the speed and accuracy of management and financial reporting without the cost and delay of operational system standardization. Kalido customers who have measured the business benefits of their projects have typically found they have derived annual savings of millions, and in some cases tens of millions, of dollars through improved management of their company performance and reduction of IT costs.

With KALIDO, companies can rapidly create and manage adaptive data warehouses and associated master data throughout their lifecycle, benefiting from the software's strategic flexibility, low cost and speed of deployment. An independently audited study shows that KALIDO saves as least 55 percent in ownership costs compared to custom-built approaches, a figure surpassed by real-life customer experiences. A typical KALIDO data warehouse implementation takes 2-5 months, as opposed to 9-18 months for conventional methods.

Kalido customers include some of the largest companies in the world, such as BP, Cadbury Schweppes, HBOS plc, Intelsat, Labatt Breweries of Canada, Owens Corning, Philips, Royal Dutch/Shell Group of Companies (Shell) and Unilever. These companies and many others use Kalido's award-winning software in over 100 countries for their enterprise-wide data warehousing and master data management projects.

A privately-held company, Kalido is headquartered in Burlington, Mass. and London, UK and has regional sales offices throughout the United States, United Kingdom, and France. More information about Kalido can be found at: http://www.kalido.com.

For more information please contact us:

www.kalido.com
info@kalido.com

www.kalido.com

Kalido
1 Wayside Road
Burlington, MA 01803
USA
Tel: +1 781 202 3200

Kalido
1 The Strand
London
WC2N 5AB
United Kingdom
Tel: +44 (0) 20 7484 2200

Version No: WP-AAII-ENG-1104

KALIDO